

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

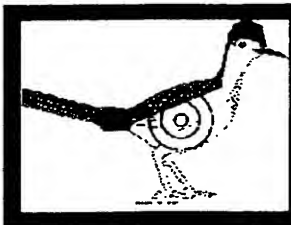
**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**




FIG. 1

1/40

File Edit View Go Communicator Help
Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

**Target Mining Interface**



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:
☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebra fish)

100%

FIG. 2A

2/40

File Edit View Go (Communicator) Help

Bookmarks Location: http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	% ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (F50)	Best Iter. (F50)	E- value (F)
J	AAAS9544.1 drill through Top50BlastHits Red Seq View	AAA59544.1	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
→	AAB24821.1 drill through Top50BlastHits Red Seq View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
J	Q99715 drill through Top50BlastHits Red Seq View	Q99715	COLLAGEN ALPHA I(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E-
J	P20701 drill through Top50BlastHits Red Seq View	P20701	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN)(CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human).	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
J	AAC31672.1 drill through Top50BlastHits Red Seq View	AAC31672.1	leukocyte function-associated molecule - 1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
J	CAA72402.1 drill through Top50BlastHits Red Seq View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E-
J	AAB38702.1 drill through Top50BlastHits Red Seq View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E-
J	CAB70853.1 drill through Top50BlastHits Red Seq View	CAB70853.1	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E-
J	CAA27972.1 drill through Top50BlastHits Red Seq View	CAA27972.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E-
J	AAB59512.1 drill through Top50BlastHits Red Seq View	AAB59512.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E-
J	CAA07569.1 drill through Top50BlastHits	CAA07569.1	matriin-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E-

FIG. 2B

1) 509 hits identified by Genome Threader only:

AddList	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	%ID	Query rgn.	Target rgn.	Aln. score	Conf.
J	Red Seq. View	AAF71133.1 drill through Top50BlastHits	AAF71133.1	PR02769	Homo sapiens	PRI	13.8% unmasked SW	109-179	1-80	122	100% unmasked GT
J	Red Seq. View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	9.6% unmasked SW	7-117	20-124	82	99.18% unmasked GT
J	Red Seq. View	CAA82910.1 drill through Top50BlastHits	CAA82910.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	11% unmasked SW	5-164	61-225	75	99.18% unmasked GT
J	Red Seq. View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	9.6% unmasked SW	7-117	318-422	82	98.86% unmasked GT
J	Red Seq. View	BAA20761.1 drill through Top50BlastHits	BAA20761.1	Not given	Homo sapiens	PRI	16.4% unmasked SW	5-114	1836-1950	78	98.52% unmasked GT
J	Red Seq. View	AAA36154.1 drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	13.2% unmasked SW	5-112	10-137	79	98.16% unmasked GT
J	Red Seq. View	AAF03046.1 drill through Top50BlastHits	AAF03046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	13.8% unmasked SW	5-113	4-131	79	97.07% unmasked GT
	Red Seq. View	AAC74854.1 drill through Top50BlastHits	AAC74854.1	orf, hypothetical protein	Escherichia coli	BCT	16.7% unmasked SW	5-89	250-335	78	95.09% unmasked GT
J	Red Seq. View	AAB60942.1 drill through Top50BlastHits	AAB60942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	16.2% unmasked SW	63-156	2-90	80	93.33% unmasked GT
J	Red Seq. View	AA467537.1 drill through Top50BlastHits	AA467537.1	glycoprotein IIIa	Homo sapiens	PRI	18.9% unmasked SW	3-112	111-251	72	92.08% unmasked GT reverse Hit
J	Red Seq. View	AA52589.1 drill through Top50BlastHits	AA52589.1	Not given	Homo sapiens	PRI	19.6% unmasked SW	3-112	138-277	74	91.34% unmasked GT reverse Hit
J	Red Seq. View	AAB71380.1 drill through Top50BlastHits	AAB71380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6% unmasked SW	3-112	138-277	74	91.34% unmasked GT reverse Hit
J	Red Seq. View	AAA35927.1 drill through Top50BlastHits	AAA35927.1	Not given	Homo sapiens	PRI	19.6% unmasked SW	3-112	138-277	74	91.34% unmasked GT

FIG. 3

File Edit Search Type Help

Redundant Sequence Display

Contains:
2 Sub-sequences.
0 PROSITE hits.
0 PRINTS hits.

Representative code: BAA15585 | Length: 427 | Organism name: Escherichia coli | [..additional annotation]

Aligned sequences	Code
	BAA15585.1
	P76235
	AAC74854.1

Primary database information

GenBank	
protein_id	Details
BAA15585.1	Tax ID: 562
EC Number: Not available	Organism name: Escherichia coli
Description: Not given	Gene Name: yz0C
Other links	
GI: 1736412	TAXON: 562

SWISS-PROT

Details	
Code	
P76235	Tax ID: 562
EC Number: Not available	Organism name: Escherichia coli
Description: HYPOTHETICAL 49.4 KDA PROTEIN IN GAP-RND INTERGENIC REGION	Gene Name: YEAH

Tools:

Sequences:

Ligands:

Start:



6/40

FIG. 4

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi

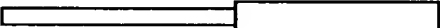

Pfam
Protein families database of alignments and HMMs
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help


Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	Align


[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families ☒ Hypertext linked to swisspfam ☐

Query gi|1788084|gb|AAC74854.1|233-423 matching Pfam-B_39416

```

YEAM_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCIMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCIMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

YEAM_ECOLI 283 TYKNVEVVYIRHQTQAKEVDENEFFYSQETG6TIVSSALKLMDEVVKERY 332
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRHQTQAKEVDENEFFYSQETG6TIVSSALKLMDEVVKERY 332

YEAM_ECOLI 333 NPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYAAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382

YEAM_ECOLI 383 QTLWREYEMLQSTFDNFAMQHIRDQDDIYPVVFRELPHKQNA 423
gi|1788084|gb|AAC74854.1| 383 QTLWREYEMLQSTFDNFAMQHIRDQDDIYPVVFRELPHKQNA 423

```

Align to family

If you think there is anything wrong with this script, please contact [Pfam](#)

100%

File Edit View Go Communicator Help	
Bookmarks Location: http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&db=	
LOCUS	AAC74854 427 aa BCT 01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].
ACCESSION	AAC74854
PID	g1788084
VERSION	AAC74854.1 GI:1788084
DBSOURCE	locus AE000273 accession AE000273.1
KEYWORDS	
SOURCE	Escherichia coli K12.
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia
REFERENCE	1 (residues 1 to 427)
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331), 1453-1474 (1997)
MEDLINE	97426517
PUBMED	9278503
REFERENCE	2 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (residues 1 to 427)
AUTHORS	Plunkett, G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIM grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names. Method: conceptual translation.
FEATURES	Location/Qualifiers
source	1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:83333"
Protein	1..427 /function="orf, Unknown" /product="orf, hypothetical protein"
CDS	1..427 /gene="yeaH" /coded_by="1788078:6385..7668" /transl_table=11 /note="o427; This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC_BAC5U SW: P45742"
ORIGIN	1 mtwfidrlrn gknksmvnrq rflrkykqi kqsiseaink rsvtdvdsge svsiptedis 61 epmfhqrggg lhrvhpqnd hfvqndrier pggggggsgs gggqasqdge gqdefvfqis 121 kdepildlfe dlalpnlkq qqrqlteykt hrqyptangv panisvvrsl qnslarrtam 181 taokrrrelha leenlaishn sepaqlleec rlrkeiaelr akiervpfid tfdlrkykne
100%	

FIG. 6A

File Edit Search Type Help

Accession Code Query

inpharmatica

Tools

Sequences

Ligands

Select a database and enter a code

Database	
GenBank	
Accession	<input type="radio"/>
Accession version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein ID	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC74854.1

Submit Reset

FIG. 6B

File Edit Search Type Help

Aligned Sequence Display

Query details:

AAC74854.1: orf, hypothetical protein

Page 1 of 2

Total hits: 36

Selected: 0

View alignment

Select all

Deselect all

inpharmatica Genome Threader results:

Cluster

Detail

Code

Title

Organism

%ID Query/Hit

Target length

Align score

Method

Confidence

1

ILFA-A

CD11A I-DOMAIN WITH BOUND MN++

Hsap|ens

16.7 250 - 335

5 - 89

78

Local

Marginal(95%)

2

1Z00A

CD11A I-DOMAIN WITH BOUND MAGNESI...

Hsap|ens

16.7 250 - 335

5 - 89

78

Local

Marginal(95%)

3

1Z00B

CD11A I-DOMAIN WITH BOUND MAGNESI...

Hsap|ens

16.7 250 - 335

5 - 89

78

Local

Marginal(95%)

4

1Z00E

CD11A I-DOMAIN WITH BOUND MAGNESI...

Hsap|ens

16.7 250 - 335

5 - 89

78

Local

Marginal(95%)

5

1Z0P-A

CD11A I-DOMAIN WITH BOUND MAGNESI...

Hsap|ens

16.7 250 - 335

5 - 89

78

Local

Marginal(95%)

6

1BHQ1

MAC-11 DOMAIN MAGNESIUM COMPLEX

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

7

1IDN1

MAC-11 DOMAIN METAL FREE

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

8

1BHQ2

MAC-11 DOMAIN CADMIUM COMPLEX

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

9

1BHQ1

MAC-11 DOMAIN CADMIUM COMPLEX

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

10

1BHQ2

MAC-11 DOMAIN MAGNESIUM COMPLEX

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

11

1IDN2

MAC-11 DOMAIN METAL FREE

Hsap|ens

15.1 250 - 355

4 - 128

76

Local

Low(94%)

12

1ATZ-B

HUMAN VON WILLEBRAND FACTOR A3 DO...

Hsap|ens

15.1 244 - 352

5 - 101

72

Local

Low(89%)

13

1ATZ-A

HUMAN VON WILLEBRAND FACTOR A3 DO...

Hsap|ens

15.3 250 - 335

5 - 89

75

Local

Low(89%)

14

1DGA-A

NMR SOLUTION STRUCTURE OF THE INSE...

Hsap|ens

16.3 250 - 331

8 - 89

75

Local

Low(89%)

15

1IDO

I-DOMAIN FROM INTEGRIN CR3 MG2+ BO...

Hsap|ens

15.1 250 - 355

4 - 128

75

Local

Low(82%)

16

1JULM

I-DOMAIN FROM INTEGRIN CR3 MIN2+ BO...

Hsap|ens

15.1 250 - 355

4 - 128

75

Local

Low(89%)

17

1ZON

CD11A I-DOMAIN WITHOUT BOUND CATI...

Hsap|ens

16.7 250 - 335

5 - 89

72

Local

Low(89%)

18

1LFA-B

CD11A I-DOMAIN WITH BOUND MN++

Hsap|ens

15.0 250 - 335

5 - 87

66

Local

Low(80%)

19

1OAK-A

CRYSTAL STRUCTURE OF THE VON WILLE...

Hsap|ens

20.3 251 - 301

1 - 59

72

Local

Low(72%)

20

1TAUC

A3 DOMAIN OF VON WILLEBRAND FACTOR

Hsap|ens

21.2 250 - 301

18 - 88

69

Local

Low(66%)

21

1AOSA

A3 DOMAIN OF VON WILLEBRAND FACTOR

Hsap|ens

15.2 245 - 331

2 - 89

63

Local

Low(53%)

22

1AOSB

A3 DOMAIN OF VON WILLEBRAND FACTOR

Hsap|ens

15.2 245 - 331

2 - 89

63

Local

Low(53%)

23

1AOSX-B

I DOMAIN FROM INTEGRIN ALPHAZ-BETA1

Hsap|ens

13.1 242 - 335

1 - 105

59

Local

Low(21%)

24

1RYPD

CRYSTAL STRUCTURE OF THE 20S PROTE...

S cerevisiae

13.1 246 - 305

166 - 226

58

Local

Low(21%)

25

1RYP-R

CRYSTAL STRUCTURE OF THE 20S PROTE...

S cerevisiae

13.1 246 - 305

166 - 226

58

Local

Low(21%)

26

1CK4B

CRYSTAL STRUCTURE OF RAT A1B1 INTEG...

Rattus norvegicus

10.5 250 - 381

5 - 86

55

Local

Low(20%)

Tools

Sequences

Ligands

GenBank

SWISS-PROT

PDB

9/40

FIG. 7

ALEye output (January 4, 2002 3:07 PM)

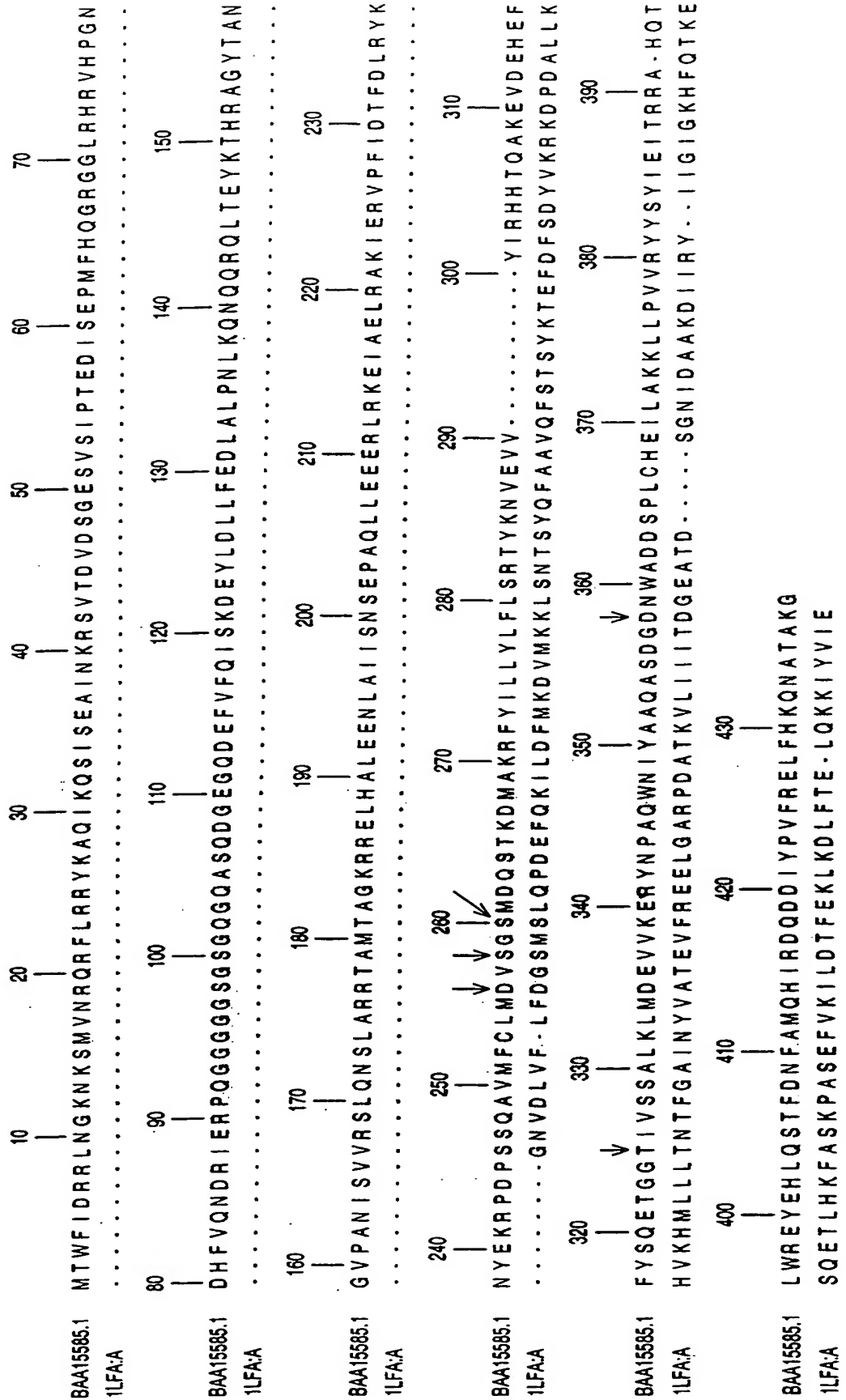
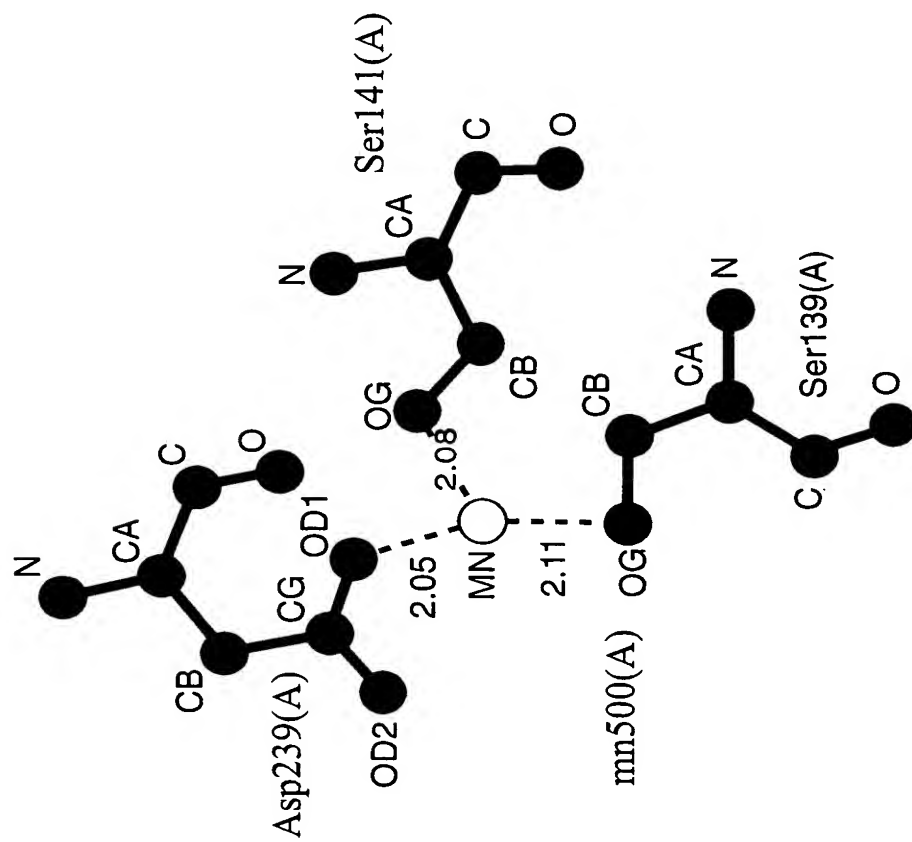


FIG. 8A



1lfa: MN500 Chain [A]

FIG. 8B

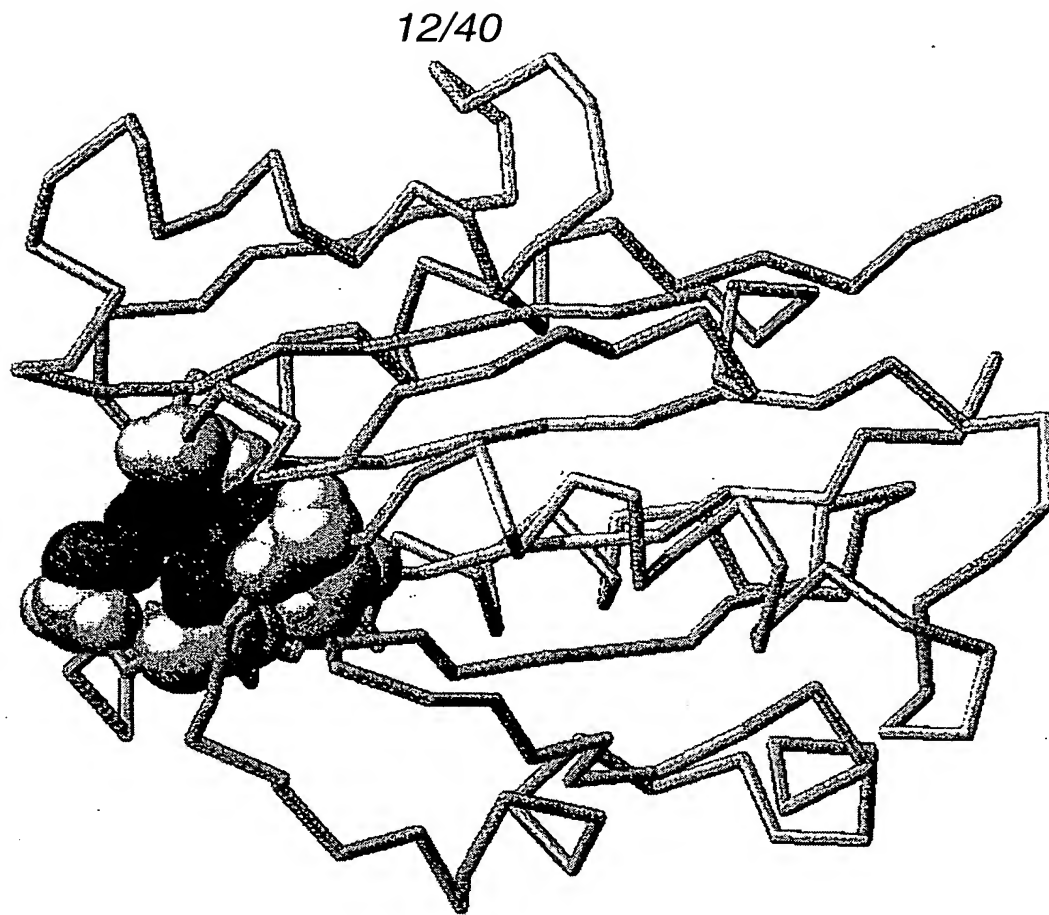
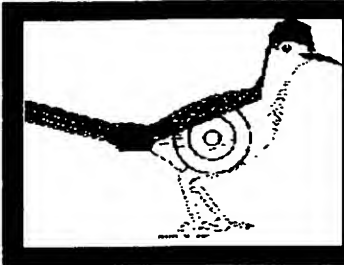



FIG. 9

13/40

File Edit View Go Communicator Help
Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIG. 10A

File

Edit

View

Go

Communicator

Back

Forward

Reload

Home

Search

NetScope

Print

Security

Shop

Bookmarks

Location

http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl#Mine2

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-BLAST output: PSI-BLAST values are shown in square!

Accession	RefSeq link	WWW link	Title	Organism	Div.	%ID (GI/PSI)	Query run (GI/PSI)	Target run (GI/PSI)	Ann. score (GI)	Conf. (GI)	1st iter. (PSI)	Best iter. (PSI)	Best E-value (PSI)
J	P55199 orf1 through 10550881.Hits Ref. Seq. View	P55199	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A)	Homo sapiens (Human)	PRI	51.8%, 51% unmasked SW	2-200, 2-200	139-337, 139-337	487	100% unmasked GI	1	2	2E-73
J	P17201 orf1 through 10550881.Hits Ref. Seq. View	P17201	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD48B)	Homo sapiens (Human)	PRI	99.5%, 100% unmasked SW	2-200, 2-200	189-387, 189-387	471	100% unmasked GI	1	1	1E-111
J	AA01258.1 orf1 through 10550881.Hits Ref. Seq. View	AA01258.1	Integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmasked SW	2-200, 2-200	159-355, 159-355	469	100% unmasked GI	1	2	4E-68
J	AA051919.2 orf1 through 10550881.Hits Ref. Seq. View	AA051919.2	Integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmasked SW	2-200, 2-200	159-355, 159-355	469	100% unmasked GI	1	2	4E-68
J	Q92715 orf1 through 10550881.Hits Ref. Seq. View	Q92715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR	Homo sapiens (Human)	PRI	28.1%, 31% unmasked SW	2-186, 7-189	133-318, 2323-2495	437	100% unmasked GI	1	2	2E-42
J	AAAS9491.1 orf1 through 10550881.Hits Ref. Seq. View	AAAS9491.1	Not given	Homo sapiens	PRI	27.1%, 27% unmasked SW	2-200, 2-200	144-338, 144-338	435	100% unmasked GI	1	2	1E-60
J	AAAS9544.1 orf1 through 10550881.Hits Ref. Seq. View	AAAS9544.1	Not given	Homo sapiens	PRI	27.1%, 27% unmasked SW	2-200, 2-200	144-338, 144-338	435	100% unmasked GI	1	2	1E-60
J	CA071722.1 orf1 through 10550881.Hits Ref. Seq. View	CA071722.1	alpha2(BSD)15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9%, 26% unmasked SW	5-195, 156-201	293-472, 1581-1626	424	100% unmasked GI	2	3	5E-6
J	P21841 orf1 through 10550881.Hits Ref. Seq. View	P21841	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1)	Homo sapiens (Human)	PRI	26.8%, 26% unmasked SW	3-198, 3-198	271-455, 37-224	424	100% unmasked GI	1	2	3E-46
J	AA001506.1 orf1 through 10550881.Hits Ref. Seq. View	AA001506.1	type XII collagen	Homo sapiens	PRI	27.6%, 30% unmasked SW	2-195, 2-195	133-318, 133-318	422	100% unmasked GI	1	2	3E-50
J	AA031952.1 orf1 through 10550881.Hits Ref. Seq. View	AA031952.1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	46.5%, 46% unmasked SW	2-189, 2-189	162-359, 162-359	419	100% unmasked GI	1	2	5E-63
J	CA007569.1 orf1 through 10550881.Hits	CA007569.1	matrin-4	Homo sapiens	PRI	26%, 28% unmasked SW	1-189, 2-200	335-529, 28-223	411	100% unmasked GI	1	2	1E-47

FIG. 10B

GT
Confidence

	Top50BlastHits Red. Seq. View	<u>BAA91707.1</u>	Not given	Homo sapiens	PRI	100% unmasked SW	100% unmasked SW	36-141	142	100% unmasked GT	3	5	2E-8
J	AAA36795.1 drill through Top50BlastHits	<u>AAA36795.1</u>	undulin 2	Homo sapiens	PRI	34.7%, 21% unmasked SW	2-46, 139-196	3-51, 37-93	121	100% unmasked GT	2	2	8E-6
J	Red. Seq. View												
J	AAD40367.1 drill through Top50BlastHits	<u>AAD40367.1</u>	calcium-activated chloride channel-2	Homo sapiens	PRI	15%, 15% unmasked SW	8-117, 8-117	312-421, 312-421	117	100% unmasked GT	5	5	1E-4
J	Red. Seq. View												
J	CAA67559.1 drill through Top50BlastHits	<u>CAA67559.1</u>	collagen VI-alpha-1 chain	Homo sapiens	PRI	21.1%, 21% unmasked SW	108-197, 108-197	7-92, 7-92	114	100% unmasked GT	3	3	5E-7
J	Red. Seq. View												
→	AAC76768.1 drill through Top50BlastHits	<u>AAC76768.1</u>	orf, hypothetical protein	Escherichia coli	BCT	13.2%, 11% unmasked SW	9-134, 7-186	267-384, 265-424	107	100% unmasked GT	-4	-5	3E-24
J	Red. Seq. View												
J	CAB43000.1 drill through Top50BlastHits	<u>CAB43000.1</u>	tellurite resistance	Escherichia coli	BCT	12.9%, 14% unmasked SW	8-192, 2-181	214-398, 208-380	102	100% unmasked GT	-6	-7	6E-22
J	Red. Seq. View												
J	AAA60114.1 drill through Top50BlastHits	<u>AAA60114.1</u>	platelet membrane glycoprotein IIb	Homo sapiens	PRI	14%, 13% unmasked SW	157-199, 157-199	145-187, 145-187	111	98.52% unmasked GT	-7	-8	8E-5
J	Red. Seq. View									reverse hit			

FIG. 11

Inpharmatica Biopendium 1.4
 File Edit Search Type Help

Redundant Sequence Display

Contains:
 2 Sub-sequences.
 0 PROSITE hits.
 0 PRINTS hits.

Representative code: P03818 Length: 427 Organism name: Escherichia coli [...] additional annotation

Code	Aligned sequences
P03818	
AAC76768.1	
AAA62097.1	

Primary database information

SWISS-PROT	
Code	Details
P03818	Tax ID: 562 Organism name: Escherichia coli EC Number: Not available Gene Name: yem Description: HYPOTHETICAL 496 KDA PROTEIN IN ASNA-KUP INTERGENIC REGION Other links ECOGENE: EGI1760 PIR: A04449

GenBank	
protein id	Details
AAC76768.1	Tax ID: 562 Organism name: Escherichia coli EC Number: Not available Gene Name: yem Description: orf1 hypothetical protein

Tools

Sequences

Ligands

Start

BP04

NCBI


Microso

4:01 PM


18/40

FIG. 12

File Edit View Go Communicator Help
Bookmarks Location http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi



Pfam
Protein families database of alignments and HMMs
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help




Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_15204	204	408	2.4e-108	Align

 [427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☐ Hypertext linked to swisspfam

Query gi|2367274|gb|AAC76768.1|204-408 matching Pfam-B_15204

```

YIEM_ECOLI 204 DILRLLPPELATLGITELYEFYRRLVEKQLITYRLNGESWREKVIERP 253
gi|2367274|gb|AAC76768.1| 204 DILRLLPPELATLGITELYEFYRRLVEKQLITYRLNGESWREKVIERP 253
YIEM_ECOLI 254 VIKDYDEQPRGPFIVCVDTSGSHG6FNEQCAKAPCLALMIRIALAENRRCT 303
gi|2367274|gb|AAC76768.1| 254 VIKDYDEQPRGPFIVCVDTSGSHG6FNEQCAKAPCLALMIRIALAENRRCT 303
YIEM_ECOLI 304 IMLPSTEIVRYELSGPQ6IEQAIRFLSQQFRG6TDLASCPRIMERLQSR 353
gi|2367274|gb|AAC76768.1| 304 IMLPSTEIVRYELSGPQ6IEQAIRFLSQQFRG6TDLASCPRIMERLQSR 353
YIEM_ECOLI 354 EYFDADRVIISDFIAQRLPDDVT SKVKELQRVHQRFHVRVMSAHGKPGI 403
gi|2367274|gb|AAC76768.1| 354 EYFDADRVIISDFIAQRLPDDVT SKVKELQRVHQRFHVRVMSAHGKPGI 403
YIEM_ECOLI 404 HRIFD 408
gi|2367274|gb|AAC76768.1| 404 HRIFD 408

```

If you think there is anything wrong with this script, please contact [Pfam](#)

File Edit View Go Communicator Help

Bookmarks Location <http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&db=prote>

LOCUS AAC76768 427 aa BCT 01-DEC-2000

DEFINITION orf, hypothetical protein [Escherichia coli K12].

ACCESSION AAC76768

PIR g2367274

VERSION AAC76768.1 GI:2367274

DBSOURCE locus AE000451 accession AE000451.1

KEYWORDS

SOURCE Escherichia coli K12.

ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (residues 1 to 427)

AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Maphew, G.P.,
Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.

TITLE The complete genome sequence of Escherichia coli K-12

JOURNAL Science 277 (5331), 1453-1474 (1997)

MEDLINE 97426617

PUBMED 9278503

REFERENCE 2 (residues 1 to 427)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

REFERENCE 3 (residues 1 to 427)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459

REFERENCE 4 (residues 1 to 427)

AUTHORS Plunkett, G. III.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants M600301 and M601428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain M61655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@ambr.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
site nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlym. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

Method: conceptual translation.

FEATURES

Source 1..427
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="M61655"
/db_xref="taxon:83333"

Protein 1..427
/function="orf; Unknown"
/product="orf, hypothetical protein"

CDS 1..427
/gene="yieM"
/coded_by="complement(2367272:5249..6532)"
/transl_table=11
/note="1427; sequence change joins ORFs yieD and yieM from
earlier version"

ORIGIN

1 mrsrlkdarv pplettevmc yqqsqllstp qfivqlpqil dlhrlnspw aeqarqlvda
61 nstilsalht lflqrwrsls ivqattingq lleeereqll sevgermtls gglepilaadn
121 ntaagrlwda sagqlkrqdy qlivkypgefl neqpelkrla eqlgrrreak siprndaqne
181 tfrtavrepa tveqvvdglq qsddilrlilp velatloite lenefurrlv ebllltmrth

100%

20/40

FIG. 14A

File Edit Search Type Help

Accession Code Query

inpharmatica

Tools

Sequences

Ligands

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC76768.1

Submit Reset

AAC76768.1: orf, hypothetical protein



Tools

Sequences

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

Annotations

FIG. 14C

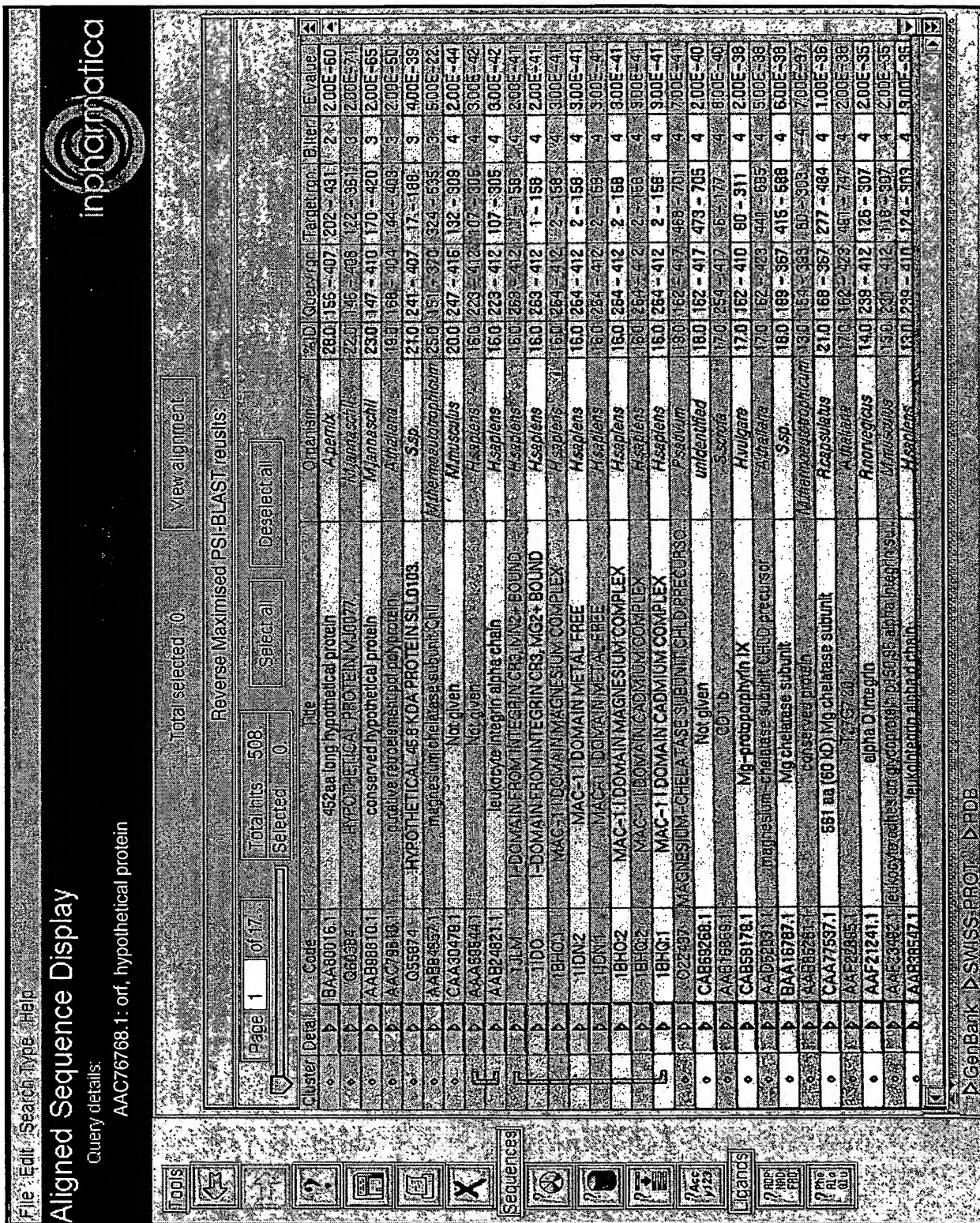


FIG. 15

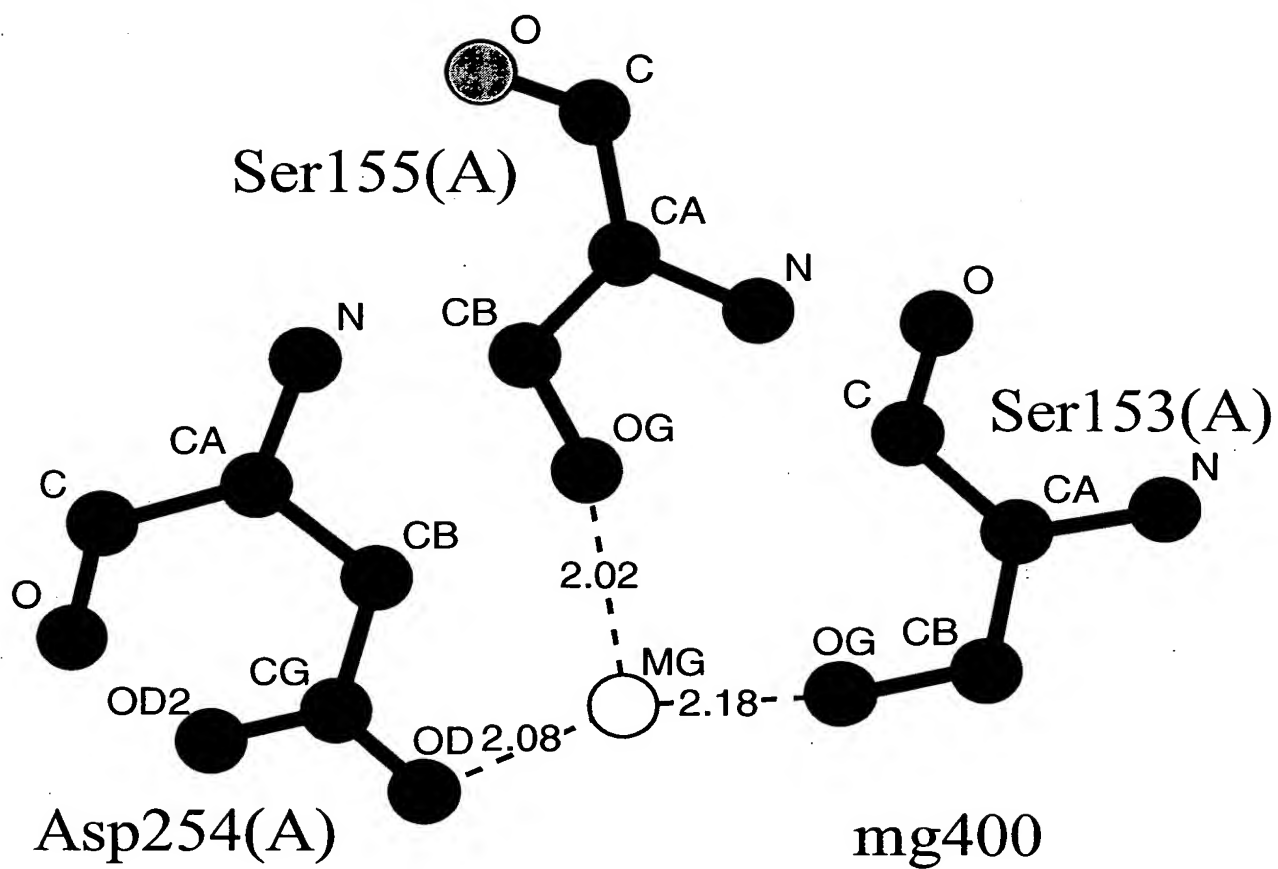
AIEye output (December 13, 2000 3:07 PM)

```

AAC78788.1      10      20      30      40      50      60      70
1A0X:A      M P S R L K D A R V P P E L T E E V M C Y Q Q S Q L L S T P Q F I V Q L P Q I L D L L H R L N S P W A E Q A R Q L V D A N S T I T S A L H T L F L Q R W R
-----
AAC78788.1      80      90      100      110      120      130      140      150
1A0X:A      L S L I V Q A T T L N Q Q L L E E E R E Q L L S E V Q E R M T L S G Q L E P I L A D N N T A A G R L W D M S A G Q L K R G D Y Q L I V K Y G E F L N E Q P
-----
AAC78788.1      160      170      180      190      200      210      220      230
1A0X:A      E L K R L A E Q L G R S R E A K S I P R N D A Q M E T F R T M V R E P A T V P E Q V D G L Q Q S D D I L R L L P P E L A T L G I T E L E Y E F Y R R L V E
-----
AAC78788.1      240      250      260      270      280      290      300
1A0X:A      K Q L L T Y R L H G E S W R E K V I E R P V V H K D Y D E Q P R G P F I V C V D T S G S M G G F N E Q C A K A F C L A L M R I -- A L A E N P R C Y I M L
-----
AAC78788.1      310      320      330      340      350      360      370      380
1A0X:A      F S T E I - V R Y E L S G P Q G -- I E Q A I R F L S Q Q F R G G T D L A S C F R A I M E R L Q S -- -- R E W F D A D A V V I S D F I A Q R L P D D
Y A N N P R V V F N L N T Y K T K E E M I V A T S Q T S Q Y G G D L N T F G A I Q Y A R K Y A Y S A A S G G P R S A T K V M V V V T D G E S H - D G S M
-----
AAC78788.1      390      400      410      420      430      440      450      460
1A0X:A      V T S K V K E L Q R V H Q H R F H A V A M S A H G K P G I M R I F D H I W R F D T G M R S R L L R R W R R
L K A V I D Q C N H D N - I L R F G I A V L G Y - -- -- -- L N R N A L D T K N L I K E I K A I A S I P T E R Y F F N V S D E A A L L E K A G
-----
AAC78788.1      470
1A0X:A      T L G E Q I F S I E G G T

```

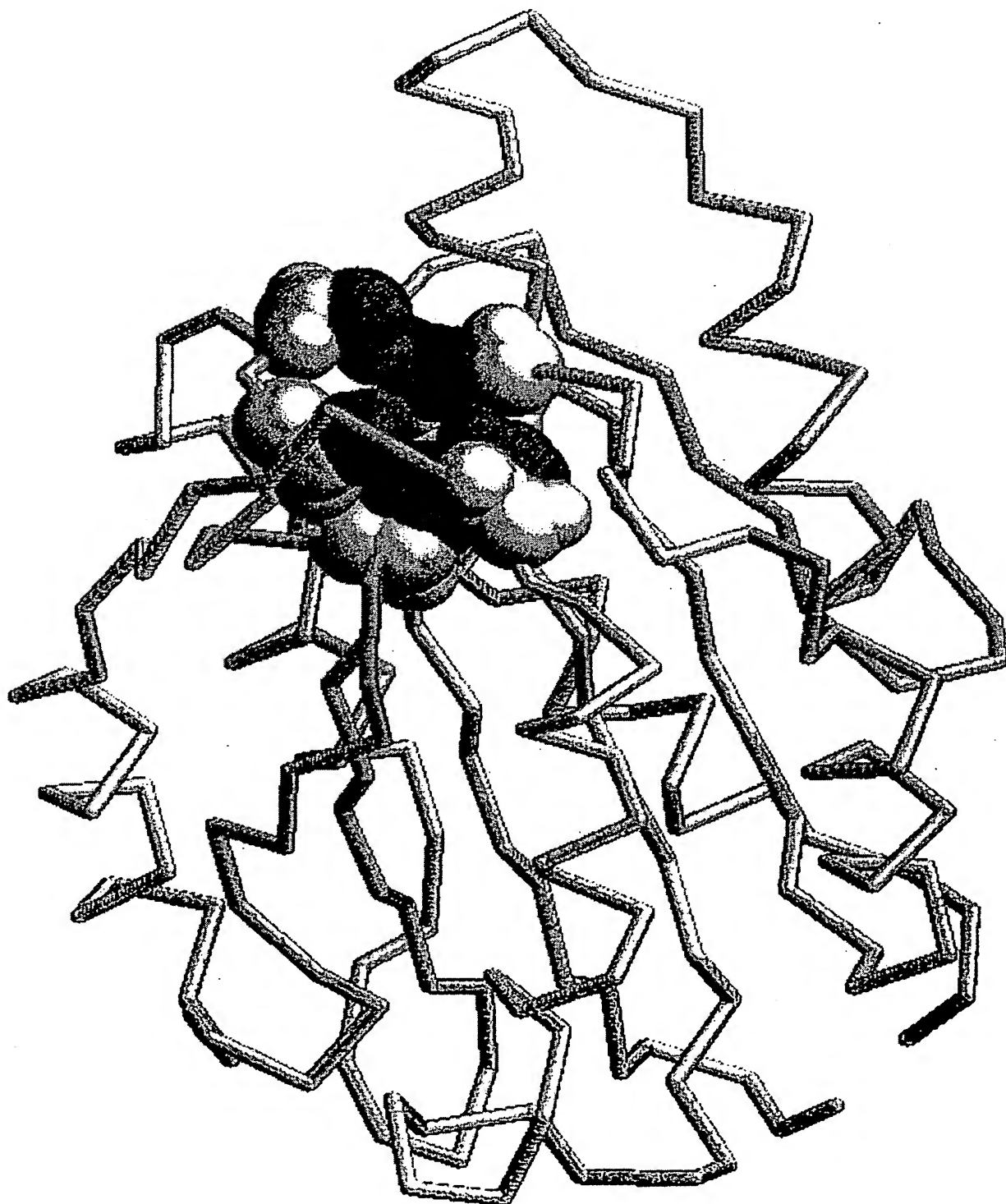
FIG. 16A



1aox: MG400

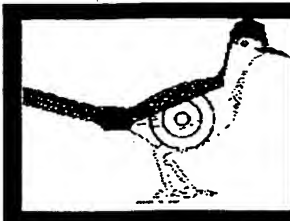
25/40

FIG. 16B




File Edit View Go Communicator Help

Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:
☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebra fish)

100%

FIG. 18A

27/40

File Edit View Go Communicator Help											
Back Forward Reload Home Search Netscape Print Security Stop											
Bookmarks Location: http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl											
2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:											
Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!											
Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st filter (PSI)
J	AAA59544.1 drill through Top50BlastHits Red Seq View	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
→	AAB24821.1 drill through Top50BlastHits Red Seq View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
J	Q99715 drill through Top50BlastHits Red Seq View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
J	AAB38702.1 drill through Top50BlastHits Red Seq View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
J	AAC01506.1 drill through Top50BlastHits Red Seq View	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
J	CAA72402.1 drill through Top50BlastHits Red Seq View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
J	AAB38547.1 drill through Top50BlastHits Red Seq View	AAB38547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
J	CAB71222.1 drill through Top50BlastHits Red Seq View	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
J	CAA07569.1 drill through Top50BlastHits Red Seq View	CAA07569.1	matrilin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
J	CAB46380.1 drill through Top50BlastHits Red Seq View	CAB46380.1	dJ453C12.3 (matrilin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

[illegible]

FIG. 19

Biopendium additional annotation page - Netscape

File Edit View Go Communication Help

Back Forward Reload Home Search Netscape Print Security Shop Help

Location: /london-bridge/inpharmatica/BPDDDEV7/cgi-bin/rsp.pl/?rep_esn=55817&passwrd=carss_app&username=carss_app&oracle_sid=BPDDDEV7

What's Related

Aligned annotation view for P10155 (downloading image...)

1: AA434452.1
Rep: P10155

1 50 100 150 200 250 300 350 400 450 500

Primary database information:

Secondary database information:

Inpharmatica calculated information:

Sequence information

Source databank	SWISSPROT
Accession code	P10155
Gene name	SSA2 OR R060
Download sequence in FASTA format	

Links to other resources:

View custom hyperlinks

Search in ExPASy

European Bioinformatics Institute	EP000504
Online Mendelian Inheritance in Man	234700
Online Mendelian Inheritance in Man	680063
Protein Information Resource	A30596
Protein Information Resource	A31760

Swissprot comments:

FUNCTION:


UNKNOWN

FOUR SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT. EACH OF

Documents Done

31/40

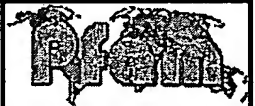
FIG. 20



Pfam

Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)



Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[338 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☐ Hypertext linked to swisspfam

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

```

92787 1 MEE SVNQMQLNEKQIANSDGCVVQVTDMMRLMRFLCF6SEG6TYYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEE SVNQMQLNEKQIANSDGCVVQVTDMMRLMRFLCF6SEG6TYYIKE 50

92787 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSF5QEGRTTKQEPMLFALAICS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSF5QEGRTTKQEPMLFALAICS 100

92787 101 QCSDISTKQARFKAIVSEUCRIPTHLPTFIQPKKDLKESMKCGMWGRALRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCSDISTKQARFKAIVSEUCRIPTHLPTFIQPKKDLKESMKCGMWGRALRK 150

92787 151 AIADWYNEKGGMALALAVTKYQRNGWSMKDLLRLSHLKPSS5EG 194
gi|133251|sp|P10155|RO60_HUMAN 151 AIADWYNEKGGMALALAVTKYQRNGWSMKDLLRLSHLKPSS5EG 194

```

Align to family

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

```

008848 195 LAIVTKYITKGWKEVMEYKEKALSVAEKLLKYLEAVEKVKRTKDDLEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVMEYKEKALSVAEKLLKYLEAVEKVKRTKDDLEV 244

008848 245 IMIIEZHQVLREMLLTNHLKSKEVVKALLQEMPLTALLRNLCKMTANSVL 294
TMT.TERM+LVREMLLTNHLKSKEVVKALLQEMPLTALLRNLCKMTANSVL

008848 295 EPGNSEVSLICEKLNEKLLKKARIMPFHVLIALETYRAGHGLRGLKWI 344
gi|133251|sp|P10155|RO60_HUMAN 295 EPGNSEVSLICEKLNEKLLKKARIMPFHVLIALETYRAGHGLRGLKWI 344

008848 345 PDKDILQALDARPYTTFTKTEPTGKRFLAVDVSAAMNQALGSVLRAST 394

```

32/40
FIG. 21

File Edit View Go Communicator Help	
Back Forward Reload Home Search Netscape Print Security Stop	
Bookmarks Location http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot	
LOCUS	RO60_HUMAN 538 aa PRI 01-FEB-1996
DEFINITION	60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJO6REN SYNDROME TYPE A ANTIGEN (SS-A)).
ACCESSION	P10155
PID	g133251
VERSION	P10155 GI:133251
DBSOURCE	swissprot: locus RO60_HUMAN, accession P10155. class: standard. created: Mar 1, 1989. sequence updated: Mar 1, 1989. annotation updated: Feb 1, 1996. xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107626 xrefs (non-sequence databases): MIM 600063, MIM 234700, PROSITE P500030
KEYWORDS	Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.
SOURCE	human.
ORGANISM	<u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (residues 1 to 538) Deutscher, S.L., Harley, J.B. and Keene, J.D. Molecular analysis of the 60-kDa human Ro ribonucleoprotein Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	89071722
REMARK	SEQUENCE FROM N.A.
AUTHORS	2 (residues 1 to 538) Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F. Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen J. Clin. Invest. 83 (4), 1284-1292 (1989)
TITLE	
JOURNAL	
MEDLINE	89198084
REMARK	SEQUENCE FROM N.A.
COMMENT	<p>This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from http://www.expasy.ch/sprot and http://www.ebi.ac.uk/sprot</p> <p>[FUNCTION] UNKNOWN. [SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN. [SUBCELLULAR LOCATION] CYTOPLASMIC. [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN. [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP). [SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.</p>
FEATURES	Location/Qualifiers
source	1..538 /organism="Homo sapiens" /db_xref="taxon:9606"
Protein	1..538 /product="60 KD RO PROTEIN"
Region	93..98 /region_name="Domain" /note="RNA-BINDING (RNP2) (BY SIMILARITY)."
Region	124..131 /region_name="Domain" /note="RNA-BINDING (RNP1) (BY SIMILARITY)."
Region	239 /region_name="Conflict" /note="K -> R (IN REF. 2)."
Region	515..538 /region_name="Conflict" /note="GMLDMCGFDTGALDVIRNFTLDMI -> ALQNTLLNKSP (IN REF. 2)."
ORIGIN	<p>1 meesvnmqgp lnekqiansq dgyvwqvtdm nrhrflcfc seggtypike qklglenaea 61 lirliedgrg cevigeiksf sqegrtrtkqe pmlfalaics qcsdistkqa afkavsevor 121 ipthlftfiq fkdldkesmk cgmwgralrk aiadwpnekq gmalalavtk ykqrngwshk 181 dlrlrlshlqp sseglaivtk yitkgwkev hlykekalsv etekilkyle avekvkrtdk 241 elevvhllee hrlvrehllt nhllskewvk allqemplita llrnlgkmta nsvlepgnse 301 vslvceklcn ekllkarih pphilialet yktghglrgk lkwrpdeeil kaldaafykt 361 fktvptgkr fillavdvsas mngrvlgsil nastvaaamc mvvtrtekds yvvaifdemv 421 pcpvttmtl qvvlmamsqi paggtdeclp miwaqktnp advfiivtdn etfaggvhp 481 ialreyrkkm dipakliverg mtsngftiad pddrgmldmc gfdtgaldvi rnfildmi</p>

FIG. 22A

File Edit Search type Help

Accession Code Query

inpharmatica

Tools

Sequences

Links

Select a database and enter a code:

Database	Field	Value
GenBank	Accession	
	Accession Version	
	GI-ND	
	protein_id	
	GI-PID	
SWISS-PROT	Accession	
	ID	
PDB	ID	

P10155

Submit Reset

inpharmatica

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJORGEN SYNDROME TYPE A ANTIGEN (SS-A)).

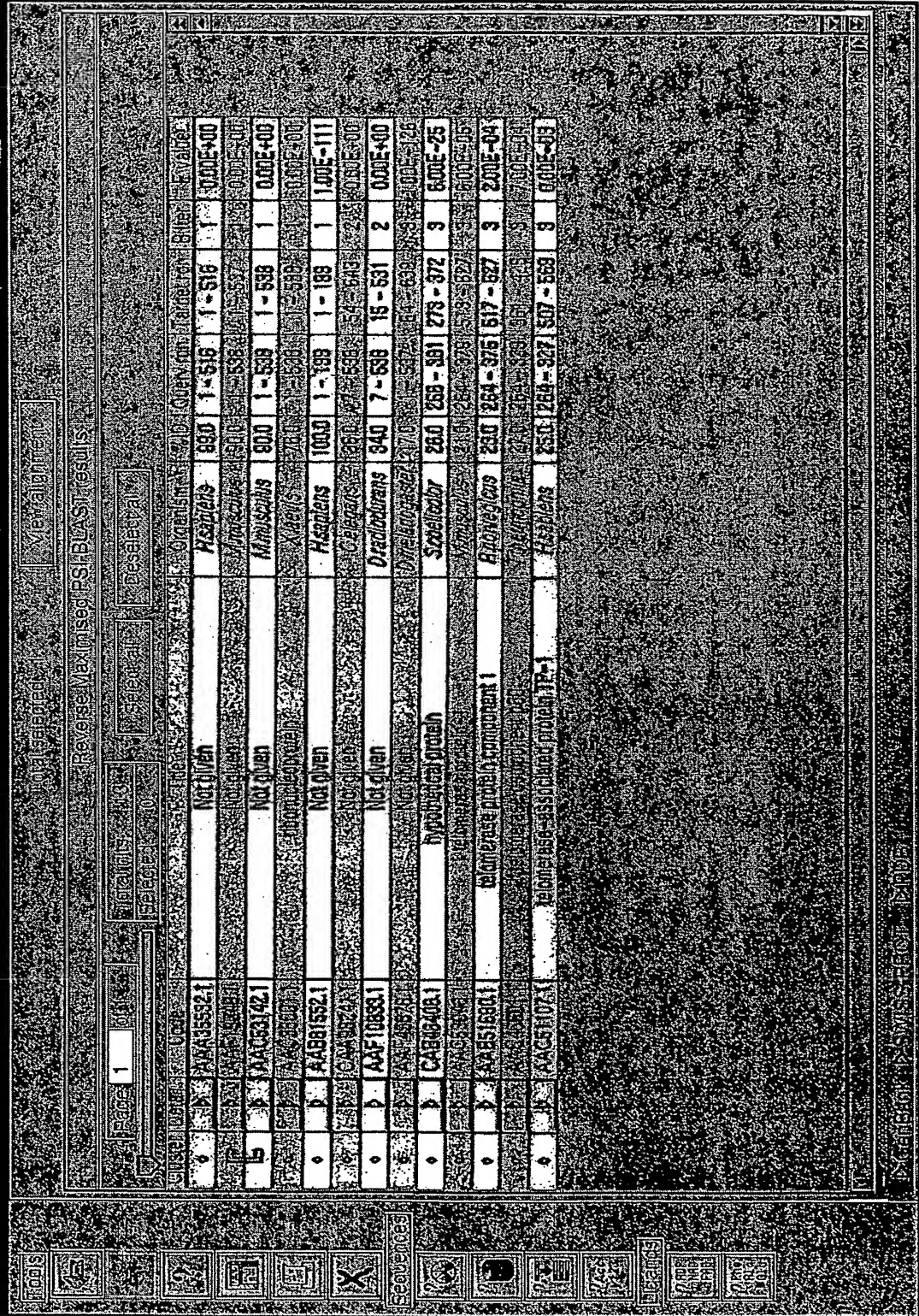


FIG. 23

AI Eye output (January 2, 2001 1:29 PM)

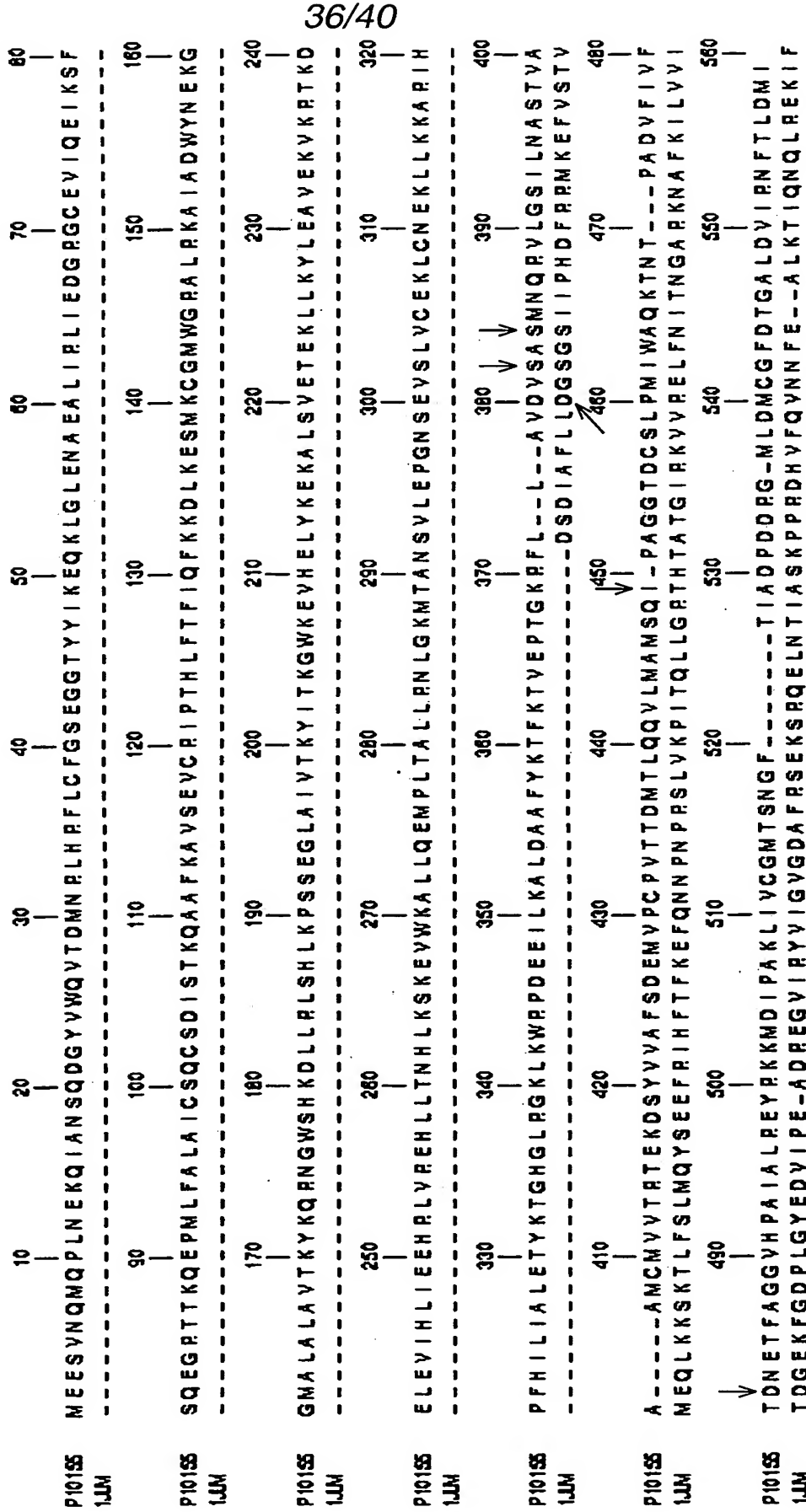
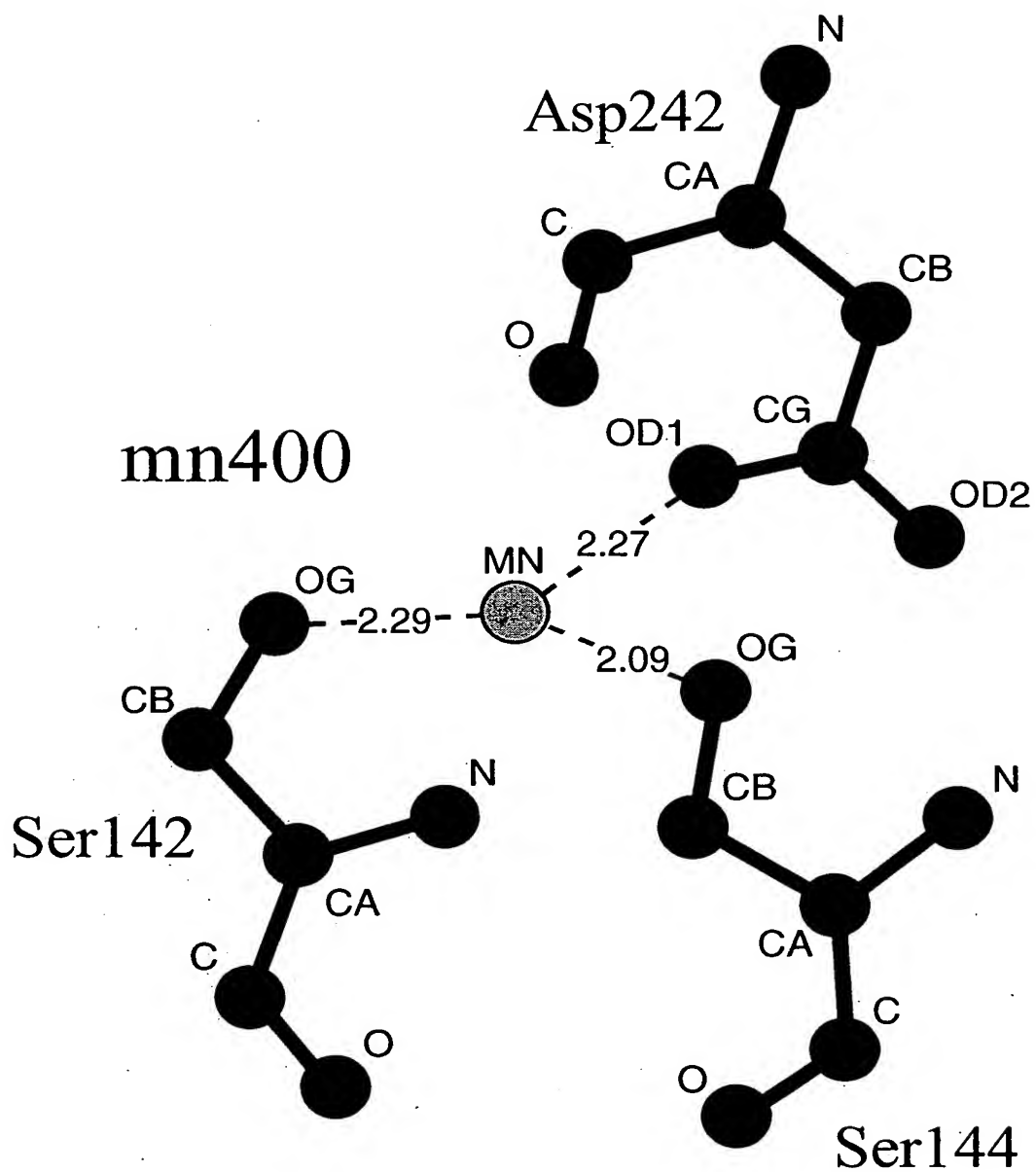


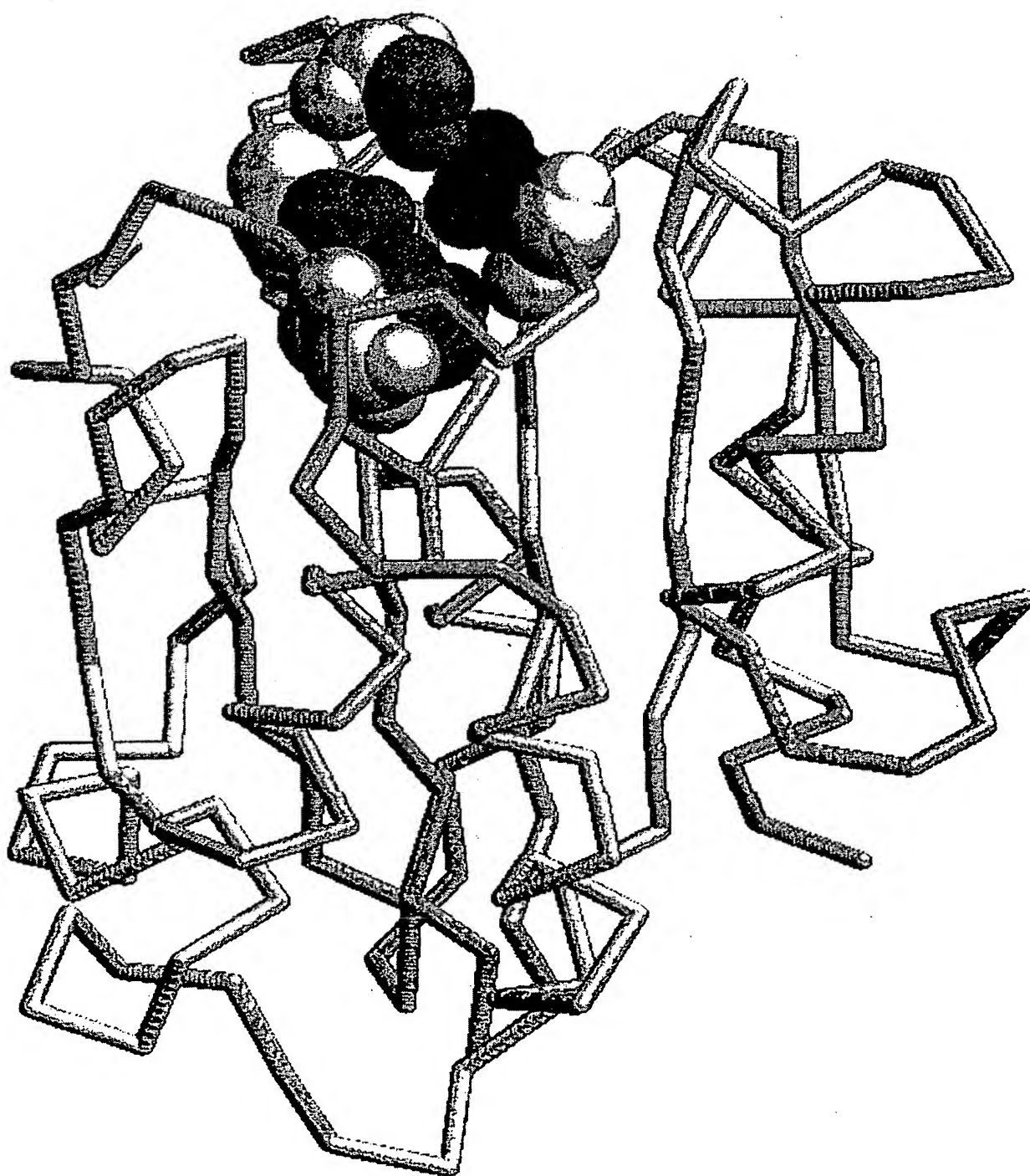
FIG. 24A



1jlm: MN 400

38/40

FIG. 24B



AIEye output (January 4, 2001 3:18 PM)

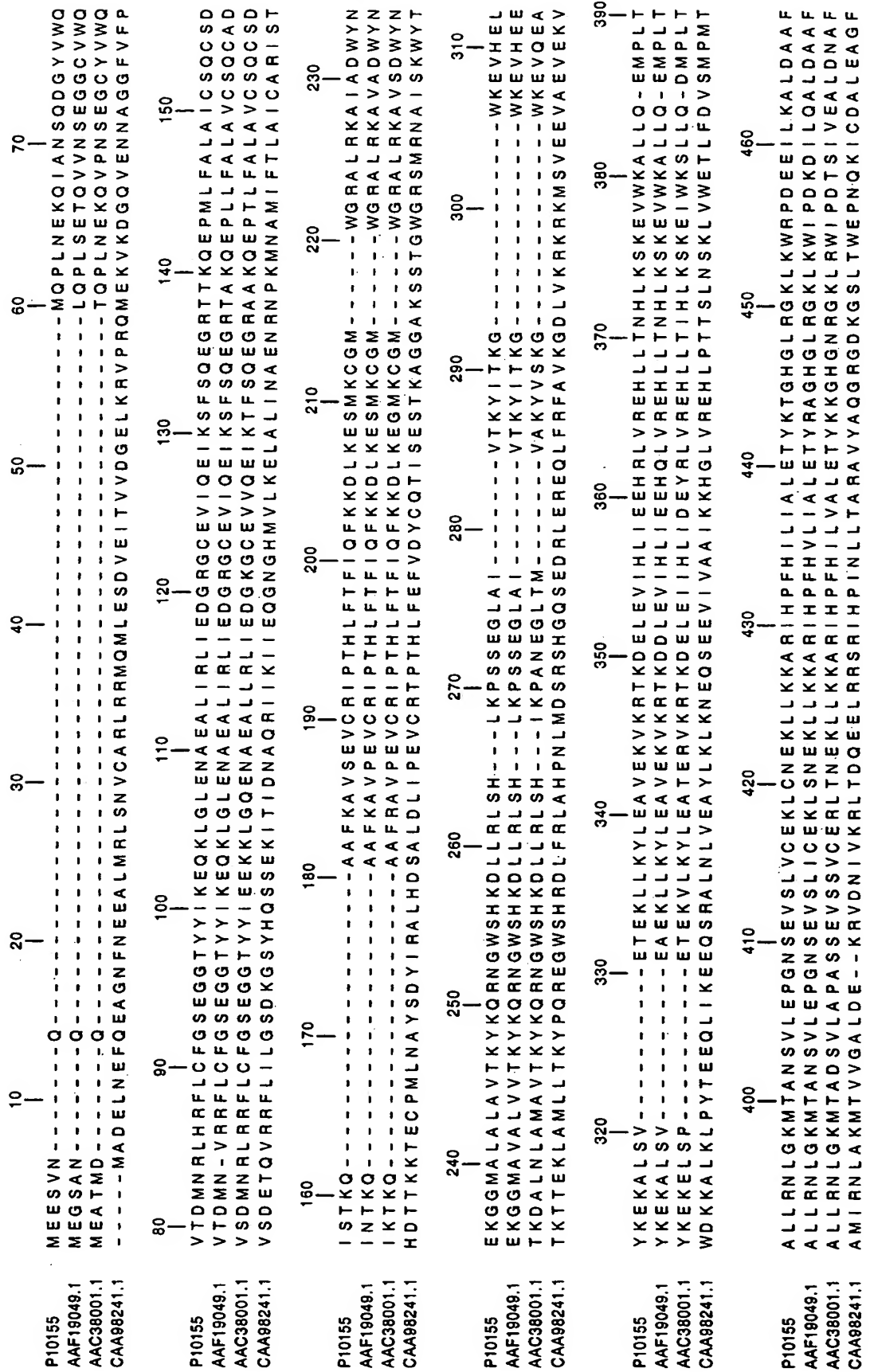


FIG. 25 (contd.)

ALEye output (January 4, 2001 3:18 PM)

[illegible]